Sparse & Low Rank — Convex Optimization for Machine Learning

Martin Jaggi

ETH Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

MADALGO CENTER FOR MASSIVE DATA ALGORITHMICS & CTIC 交互计算 Summer School 2011

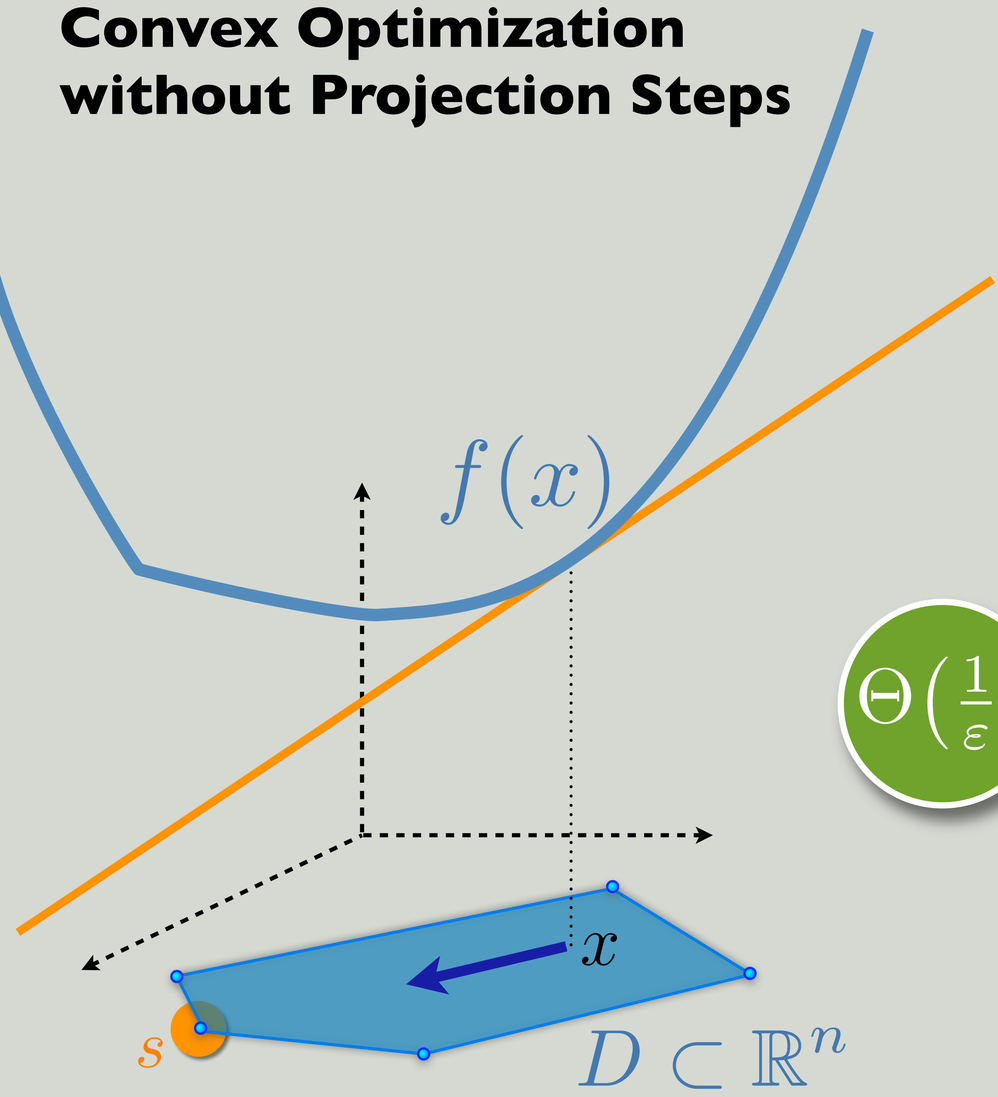
Goal:
Minimize a convex function f over a compact convex domain $D \subset \mathbb{R}^n$

Algorithm:
In each step, we greedily move towards the point in D that looks most promising as given by the current linearization

Algorithm 1 Greedy on a Compact Convex Set

Input: Convex function f, convex set D, target accuracy ε
Output: ε-approximate solution for problem $\min_{x \in D} f(x)$
Pick an arbitrary starting point $x^{(0)} \in D$
for $k = 0 \dots \infty$ **do**
  Let $d_x \in \partial f(x^{(k)})$ be a subgradient to f at $x^{(k)}$
  Let $\alpha := \frac{2}{k+2}$
  Compute $s := \operatorname{approx} \arg\min_{y \in D} y^T d_x$
    {Approximate the linearized primitive problem}
  Update $x^{(k+1)} := x^{(k)} + \alpha(s - x^{(k)})$
end for

Convex Optimization without Projection Steps



Convergence Analysis:
Algorithm obtains an ε-approximate solution in $O\left(\frac{1}{\varepsilon}\right)$ many iterations.
[J arXiv Aug 2011]

Sparse Solutions for Vector Problems

$D = \begin{cases} \text{unit simplex in } \mathbb{R}^n \\ \ell_1\text{-ball} \end{cases}$

$\Rightarrow$ ε-approximate solutions of **sparsity** $O\left(\frac{1}{\varepsilon}\right)$ [Clarkson SODA '08] (and this is best possible)

$\Theta\left(\frac{1}{\varepsilon}\right)$

Low-Rank Solutions for Matrix Problems

$D =$ symmetric matrices of unit trace

$\Rightarrow$ ε-approximate solutions of **rank** $O\left(\frac{1}{\varepsilon}\right)$ [Hazan LATIN '08] (and this is best possible)

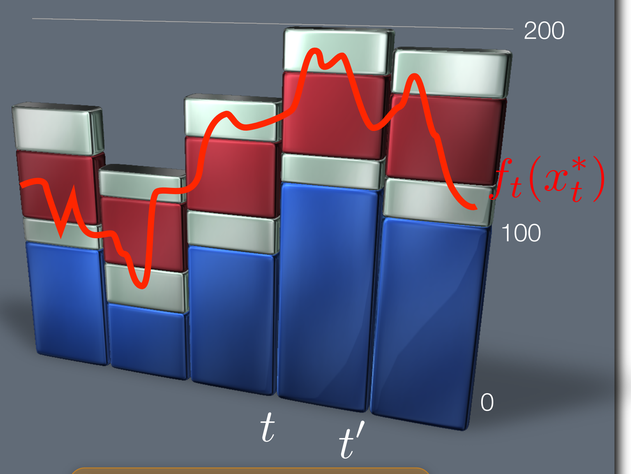
Pathwise Optimization

Parameterized Problems

$$\min_{x \in S_n} f_t(x)$$

Want: guarantee on the duality gap along the entire path,

$$g_{t'}(x) \leq \varepsilon$$



Theorem
There are $O\left(\frac{1}{\varepsilon}\right)$ many intervals of piecewise constant ε-approx. solutions.

[GJL ESA 2010]

Applications for Vector Problems

Smallest Enclosing Ball — Coresets of size $\lceil \frac{1}{\epsilon} \rceil$

Polytope Distance — Coresets of size $\lceil \frac{E}{\epsilon} \rceil$

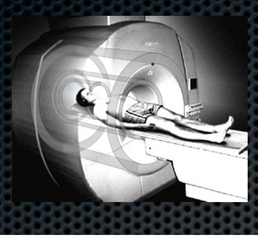
[GJ SoCG 2009]

Classification
Support Vector Machine (SVM)
Perceptron

Finding the optimal hyperplane (of maximum margin)

Regularized Regression
ℓ_1-reg. Regression
SVR
Logistic Regression

Compressed Sensing



recovery of	*minimize*
sparse vectors:	or ℓ_1 -norm / ℓ_∞ -norm
low-rank matrices:	*trace*-norm or *max*-norm

Applications for Matrix Problems

Trace-Norm Optimization

(or also for the matrix max-norm)

$$\min_{X \in \mathbb{R}^{n \times m}} f(X) + \mu \|X\|_*$$

arbitrary convex function on matrices

also called „trace norm", sum of singular values

[JS ICML 2010]

Semidefinite Optimization

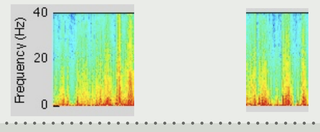
bounded **trace**-norm	bounded **max**-norm
$\min_X \; f(X)$	$\min_X \; f(X)$
$s.t. \;\; Tr(X) \leq t$	$s.t. \;\; X_{ii} \leq t \quad \forall i$
$X \succeq 0$	$X \succeq 0$

or also general semidefinite programs (SDPs)

Various Applications

Avalanche detection from Audio Data

noise/wind — avalanche



Finance: Mean-Variance Portfolio Analysis

$$\max_x \; -x^T A A^T x + \mu \cdot b^T x$$

Risk (Covariance) — Expected return

Text Classification and Similarity Measures

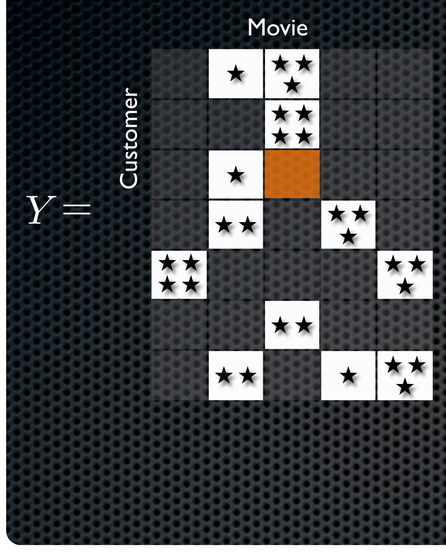
automatic quality ranking of YouTube comments

good	bad
ctually has meaning. i love co	this guy is a dushhhhhhhh...
g in my head. and it made me	wned... its a fuking diseasee
d by the media. that's why i d	tink he got fukd in the ass b
what he sang there was good.	your cats suck. i hope they
s. but i think i won't just to	this is the unfunniest shit i'
ren!. danke im voraus !. geile	its a stupid song
le for someone you love that	wat is this crap lol crap crap
thank you for uploading.. that	fake!!!!!!!!!!!!!!!!

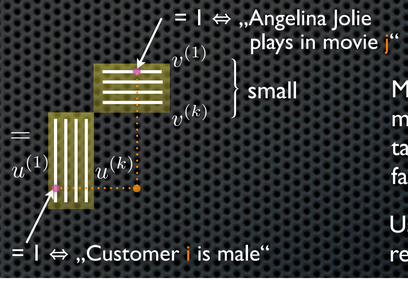
Matrix factorizations for recommender systems

The Netflix challenge:
17'000 Movies
500'000 Customers
100'000'000 Ratings
(Observed Entries ≈ 1%)

Given a small sample of entries of a matrix, we want to predict all its entries.



$\approx UV^T$



Most popular methods for this task use matrix factorizations,

Use a trace norm regularization!

Dimensionality Reduction

Principal Component Analysis
Sparse PCA,
Robust PCA